Multiple Model Tree Meta Algorithms Improvement of Network Intrusion Detection Predictions Accuracy

[1]Olasehinde Olayemi O., [2]Olayemi Olufunke C., [3]Alese B. K.
[1]*Federal Polytechnic,* [2]*Joseph Ayo Babalola University,* [3]*Federal University of Technolog*
Nigeria

Abstract

Security of Information is a critical issue for many organizations. Intrusion Detection systems (IDSs) protect information system by analyzing network packet to determine if it is abnormal or normal. This paper applies Multiple Model Trees (MMT) stacked ensemble algorithm to improve the classification accuracy of network intrusion. The predictions of the K Nearest Neighbor, Decision Tree and Naïve Bayes intrusion detection models built with UNSW-NB15 intrusion detection training dataset served as input to Multiple Model Tree (MMT)meta learner algorithm via a ten-fold cross validation to build the MMT stacked ensemble model used for the final binary classifications of the network traffics (attacks and normal) and multi-class classification into any of the nine network attacks or normal. The evaluation of all models on the testing dataset results show that MMT algorithm improves the prediction accuracy of each of the three base machine learning model predictions, It recorded the highest classification accuracy of 97.93% and lowest false alarm rate of 0.22% for the binary classification and improves the multi-class classification accuracy of all the base models prediction

1. Introduction

The rapid expansion of computer usage and computer networks has made the protection of data and computer resources against unauthorized access a big challenge to the cyber community, organization data are valuable asset and have to be protected and be made inaccessible to any unauthorized parties [1], the integrity, availability and confidentiality of such data has to be preserved against, unauthorized access, modification, usage, disclosure, theft and destruction. A system can only be considered secure if the three principles of computer security, Confidentiality, Integrity and Availability (CIA) are successfully satisfied [2]. Data mining are automatic techniques for discovering hidden and potentially useful patterns from a dataset and generate new information from it [3]. In the case of intrusion detection, it is used to build an intrusion detection model that distinguishes intrusive or anomalies patterns from the normal network traffic patterns. Building intrusion detection system involves the training of Machine Learning Algorithms with UNSW-NB15 intrusion dataset and its evaluation with test dataset. The sophistication of today's cyber threats makes it difficult to be detected by security tools such as access control, authentication, firewall and antivirus alone, in order to enhance the overall security of a computer network against cyber-attacks, Intrusion Detection System (IDS) has to be added to compliment the other security tools. IDSs are security mechanisms that monitor and detect intrusions on the computer systems, it is often being deployed as a second line of protection for the information systems. One of the several reasons that make intrusion detection a necessary part of the entire defense system is the short comings of several traditional systems and applications that are developed without giving consideration to the security concerns of the environment where the systems are deplored. A secured isolated system can become vulnerable when connected to the Internet. IDS identify attempt to compromise such security lapses in the software design. Another reason is the limitations of information security and software engineering practices, flaws or bugs that arise as a result of initial system design fault could be capitalized upon by cyber attackers to compromise the security of the systems or applications. IDS unlike firewall can detect modem attack environments and are able to analyze network packets, because of these reasons, Intrusion Detection System (IDSs) is designed to achieve high protection for the cyber security infrastructure [4]. Ensemble Learning are machine learning techniques that improve machine learning model predictions by combining the predictions of several models. Numerous studies have shown that it is capable of improving the combined model's performance over the best of a single model [5], it combines several machine learning techniques into one predictive model. Stacked Ensemble is a heterogeneous, parallel Meta Learning approach for improving intrusion detection models by combining predictions of several intrusion detection models known as base level models.

Veracity of malicious network activities that violate network policy and cause damage to information system has made intrusion detection systems (IDSs) an ingredient part of network security

[6]. Intrusion detection is a process of monitoring and analyzing network events for sign of intrusion [7]. Misuse also known as signature based and Anomaly are the two approaches used for IDS. The misuse detection system uses patterns of already known and stored pattern to match and identify intrusions. It compares the patterns known attacks with the captured network traffic, if there is a match between them, it generates an alert for a detection, this IDS approach accurately detects known attacks and has high level of false alarm rate with unknown attacks, this makes it impossible for it to detect new intrusions or zero-day attacks [8]. The anomaly detection generates a benchmark for normal behavior and then measures the behavior of incoming network packet with the benchmark, and any behavior that deviates outside the normal benchmark are term intrusive [9], it does not require prior knowledge of an intrusion and thus can detect new intrusions. According to [10]**,** IDS are used to protect computer information and resources against cyber attackers. It analyzes and predicts the behaviors of users as either normal and intrusion based on the learnt behaviors of the network.

2. Related Works

Projected a lightweight IDS model. Information Gain and Chi-Square approach were used to extract important features while Classic Maximum Entropy (ME) model was used to learn and detect intrusions [11]. [12] builds different machine learning models and evaluated their performances in terms of false alarm rate (FAR) and classification accuracy on the UNSW-15 dataset, the results shows that the Decision Tree model records the highest accuracy of 85.56% and the lowest FAR at 15.78%. In [13], Support vector machine (SVM) model was used to detect network intrusions using MATLAB. KDD dataset was used as a bench mark dataset for intrusions detections. The authors reported that SVM algorithm required long training time and as a result its usability is not feasible. [14] proposed a real time anomaly intrusion detection framework based on Fuzzy-Bayesian, combination of fuzzy and Bayesian classifiers was used to improve the overall performance of Bayes based intrusion detection system (IDS), KDD intrusion detection dataset is used to evaluate the performance of the framework. The results in [15] showed that the C4.5 decision tree intrusion detection model was more feasible and effective, returning a high accuracy rate of almost 90% of classification accuracy than other models used.

Stacked Ensemble is a two-level supervised, parallel learning approach for improving intrusion detection model prediction by combining predictions of diver's base-level models prediction using supervised meta level algorithm. The base-level models were built by training several machine learning algorithms with the intrusion detection dataset, the meta-level algorithms are trained with the predictions of the base-models. The result in [17] shows that stacking with Meta Decision Tree (MDTs) performs better than voting and stacking with decision trees, as well as boosting and bagging of decision trees, it furthers shows that MDTs performed slightly better than SCANN and selecting the best classifier with cross validation (Select Best), [18] investigated classification via regression, and reported that classification via Multi Response Model Trees (MMT) performs extremely better than multi response linear regression (MLR) and better than C5.0. This indicates that multi response model trees (MMT) are a very suitable choice for learning at the meta-level. Choudhury and Bhowal builds several Boosting Ensemble for intrusion detection using of many Machine Learning Algorithms, and concluded that Random forest and Bayes Net are the two most suitable algorithms in terms of classification accuracy to build Intrusion Detection models [19]. [20] Proposed a Particle Swarm Optimization (PSO) for feature selection for an ensemble of three base classifiers; (Classification and Regression Tree - CART, Random Forest- RF and C4.5 Decision tree), the implementation of ensemble system showed a promising accuracy and lower false alarm rate than existing ensemble techniques. [21] compares the classification accuracy and false alarm rate performance improvement of bagging, boosting, and stacking approaches to the ensemble of intrusion detection models, Four base algorithms; Naïve Bayes, Decision tree, JRip (rule induction), and K-nearest neighbor was used to build the bagging and boosting ensembles, additionally, each of the four base models was used in turn to combine the predictions of the rest of the base-models, the stacked ensemble approach achieves the highest classification accuracy of more that 99% for known attacks and highest accuracy of 60% for unknown attacks than the bagging and boosting approach.

3. Intrusion Detection Dataset

The University of New South Wales, Network Benchmark, 2015 (UNSW-NB 15) Intrusion detection dataset is the latest intrusion detection dataset, IXIA Perfect Storm tool was used created it in the Cyber Range Laboratory of the Australian Centre for Cyber Security (ACCS) in 2015 for research purposes in intrusion detection security domain. It is a hybrid of the realistic modern normal activities and the synthetic contemporary attack behaviors from network traffics. its training datasets contain 82, 332 records while the testing dataset contain 175, 341 records as shown in Table 1, contains nine attacks types namely; Reconnaissance, Backdrop, Analysis, Shellcode, Dos, Fuzzers, Exploits, Generic, and Worms, it has some advantages over the NSLKDD data set, the distribution of the training and testing sets is similar and its suitability to evaluate existing and new attacks in an effective and reliable manner [9].

4. Methodology

The System Architecture of the Stacked Ensemble of Intrusion Detection Systems with Multiple Model Tree Meta Algorithm is shown in Figure 1, the model building phase consists of three different stages; the dataset pre-processing stage involves the discretization of the UNSW-NB15 training dataset using supervised Class Attribute Interdependent Maximizations discretization algorithm, and the feature selection of relevant attributes. In the second stage, the pre-processed training dataset was used for the training and building of the three base models; K Nearest Neighbor, Decision tree and Naïve Bayes models and their evaluation via ten folds cross validation. In the last stage, the predictions of the base-models were used to train the Multiple Model Tree (MMT) meta algorithm and build the Stacked Ensemble Model. In the second phase, the test dataset is preprocessed and used to evaluate the base-models, their predictions were used to evaluate the stacked ensemble model to obtain an improved intrusion detection prediction. The system was implemented using R Programming languages for preprocessing of the dataset and Python for the building and evaluation of the intrusion detection models, on a Corel i7, 64bits, 3.3 GHz processor, 16MB Cache, 512GB SDD, Ms. Windows 10 server.

4.1. Stacked Ensemble with Multiple Model Trees (MMT)

Multiple Model Trees (MMT) are a form of decision tree with linear regression functions at the leaves. It converts predictions problem into a function approximation problem M5' inducer, MMT predictive accuracy is excellent with numeric attributes. Figure 2, shows the algorithm of MMT while Figure 3 shows the pictorial representation of the operations of MMT stacking, from the Figure 3, MMT first generated a derived dataset based on the distinct class label from the main UNSW-NB15 dataset, trees generated from the derived dataset are induced with M5' linear regression algorithm to create a classification function for each attack categories, a new instance to be classified were plug into each of the linear regression function, the function with the highest value is returned and the instance is classified as the attack of the function.

Table 1. Number of Normal and Attacks Connections in both the Training and Testing Dataset

	Training		Testing	
Names of Attack	**No of Connection**	**Percentage Distribution (%)**	**No of Connection**	**Percentage Distribution (%)**
Reconnaissance	3496	4.25	10491	5.98
Dos	4089	4.9	12264	6.99
Exploit	11132	13.52	33393	19.04
Shellcode	378	0.46	1133	0.65
Fuzzers	6062	7.36	18184	10.37
Backdoor	583	0.71	1746	1.00
Analysis	672	0.82	2000	1.14
Generic	18871	22.92	40000	22.81
Worms	44	0.05	130	0.07
Total No of Attacks	45332	55.06	119341	68.06
Normal	37000	44.94	56000	31.94
Total No of Connections	82332	100.00	175341	100.00

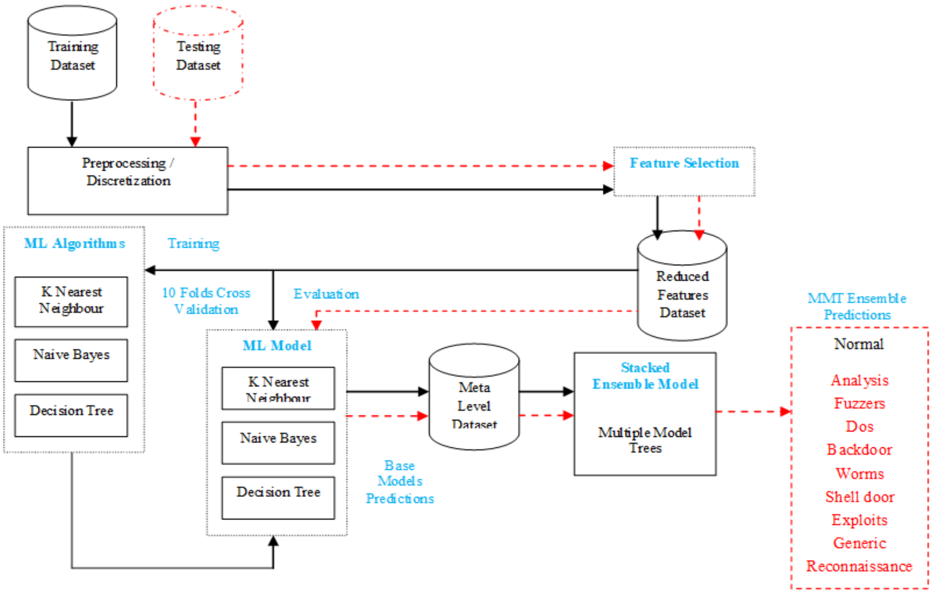
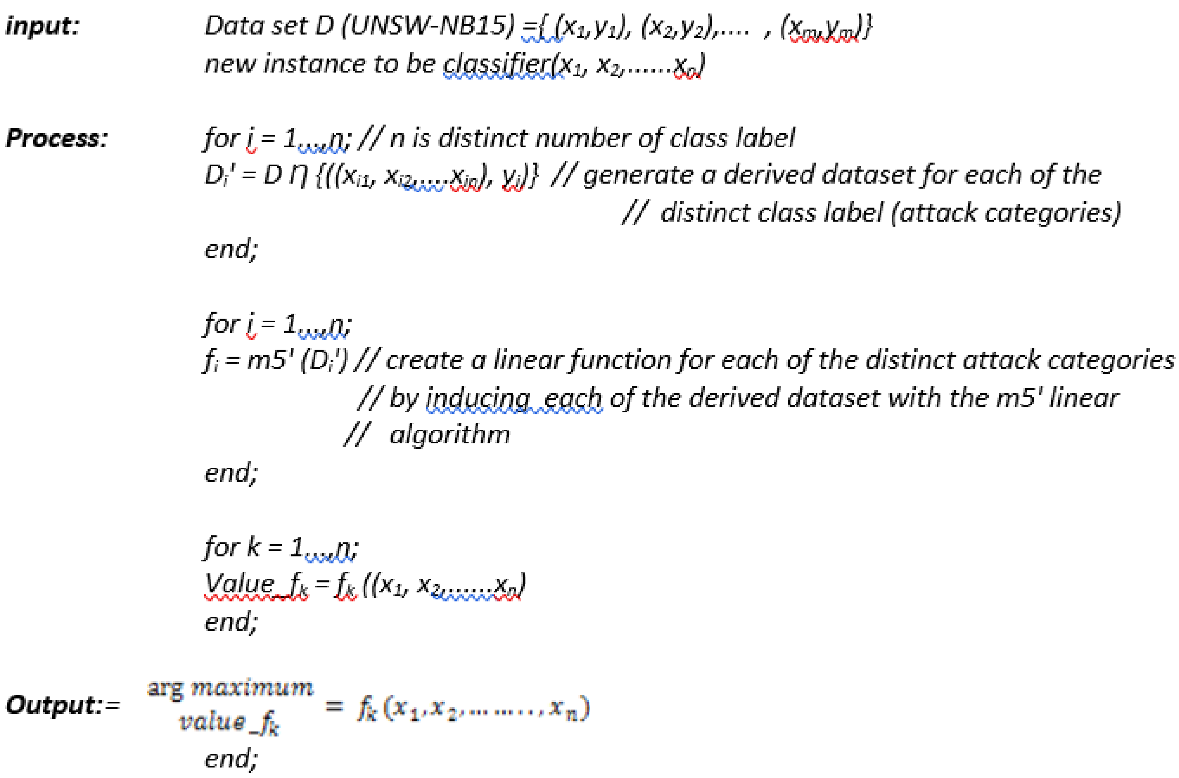
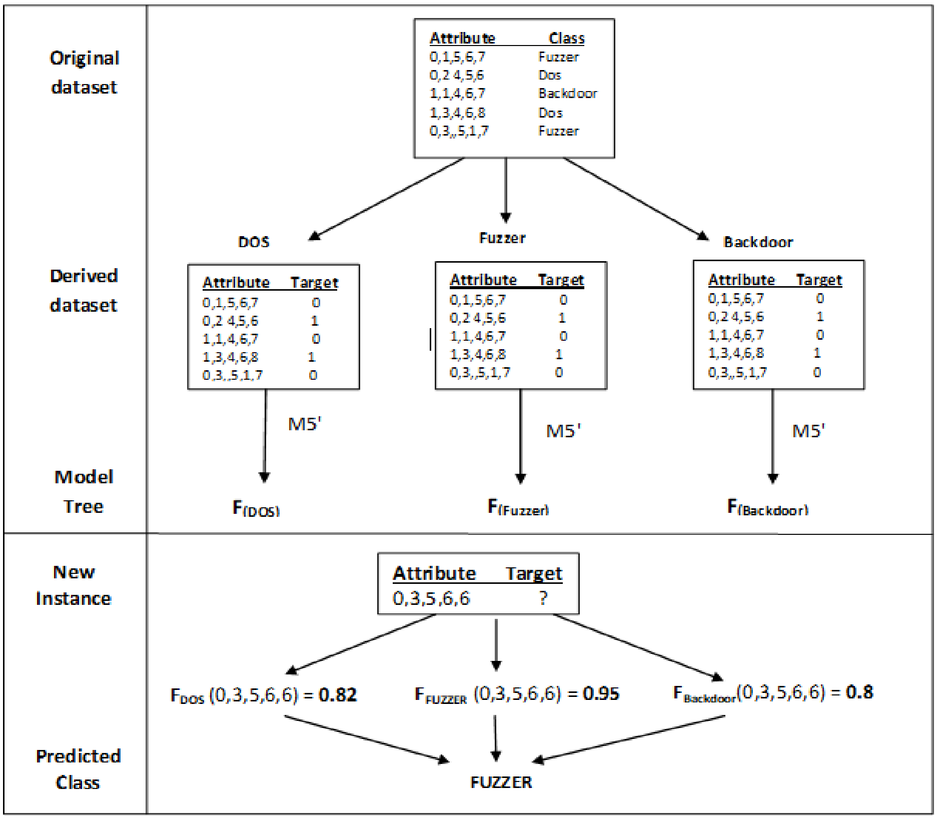


Figure 1. The Architecture of the Stacked Ensemble Network Intrusion Detection System



```
input:      Data set D (UNSW-NB15) ={ $(x_1,y_1)$, $(x_2,y_2)$,.... , $(x_m,y_m)$}
            new instance to be classifier($x_1$, $x_2$,......$x_n$)

Process:    for i = 1,...,n; // n is distinct number of class label
            $D_i$' = D ∩ {(($x_{i1}$, $x_{i2}$,....$x_{in}$), $y_i$)} // generate a derived dataset for each of the
                                            // distinct class label (attack categories)
            end;

            for i = 1,...,n;
            $f_i$ = m5' ($D_i$') // create a linear function for each of the distinct attack categories
                    // by inducing each of the derived dataset with the m5' linear
                    // algorithm
            end;

            for k = 1,...,n;
            Value_$f_k$ = $f_k$ (($x_1$, $x_2$,......$x_n$)
            end;

Output:=    arg maximum value_$f_k$ = $f_k(x_1, x_2, \dots \dots, x_n)$
            end;
```

Figure 2. Multiple Model Trees algorithm



Attribute	Class
0,1,5,6,7	Fuzzer
0,2 4,5,6	Dos
1,1,4,6,7	Backdoor
1,3,4,6,8	Dos
0,3,,5,1,7	Fuzzer

Attribute	Target
0,1,5,6,7	0
0,2 4,5,6	1
1,1,4,6,7	0
1,3,4,6,8	1
0,3,.5,1,7	0

Attribute	Target
0,1,5,6,7	0
0,2 4,5,6	1
1,1,4,6,7	0
1,3,4,6,8	1
0,3,,5,1,7	0

Attribute	Target
0,1,5,6,7	0
0,2 4,5,6	1
1,1,4,6,7	0
1,3,4,6,8	1
0,3,,5,1,7	0

Attribute	Target
0,3,5,6,6	?

Figure 3. Pictorial Representation of MMT Operation

5. Results and Discussion

The predictions of the three base models are learned (stacked) with Multiple Model Tree (MMT) algorithm to build the Stacked Ensemble model. Table 2 shows the multi-class classification accuracy of the three base models and the stacked ensemble model. MMT stacked model has the highest classification accuracy of 99.46% on Normal network connection categories, and least classification accuracy of 86.15% on Worms network connections categories, the base model's performance shows. Decision Tree model has the highest classification accuracy in seven out the ten network connection categories, followed by Naive Bayes model with the remaining three out of the ten network connection categories, despite the fact that KNN model has the least classification accuracy among the three base models, it records a higher classification accuracy than Naive Bayes in six out of the ten network connection categories, the result in Table2 further confirms, the ability of ensemble model to outperform base models, the MMT stacked Ensemble model recorded higher classification accuracy more than the highest accuracy recorded by any of the three base models.

Table 2. Multi-Class Classification Accuracy of the Prediction of the Base Models and the Ensemble Model

Network Connection Categories	Base Models			Stacked Ensemble Model
	Naive Bayes	K Nearest Neighbour	Decision Tree	
Analysis	67.35%	12.00%	22.85%	88.35%
Backdoor	61.00%	48.63%	65.41%	89.40%
Dos	66.46%	72.81%	82.11%	91.61%
Exploits	50.92%	62.94%	71.90%	89.30%
Fuzzers	37.02%	71.74%	79.42%	95.21%
Generic	95.26%	98.30%	98.60%	99.44%
Normal	67.27%	96.08%	97.38%	99.46%
Reconnaissance	55.71%	58.95%	75.19%	86.47%
Shellcode	89.41%	58.16%	69.20%	94.00%
Worms	72.31%	44.52%	67.69%	86.15%

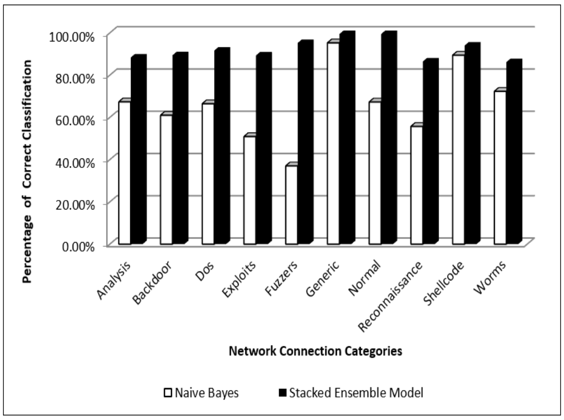


Figure 4. Classification Performance of Naive Bayes and Stacked Ensemble Model

Figures 4, 5, 6 and 7 show the graphical representations of the classification accuracy of the base and ensemble models. Table 3 shows the binary (attacks and normal label) performance of the base models and the MMT stacked model, MMT stacked model has the highest classification accuracy of 97.93% and lowest false alarm rate of 0.22%.

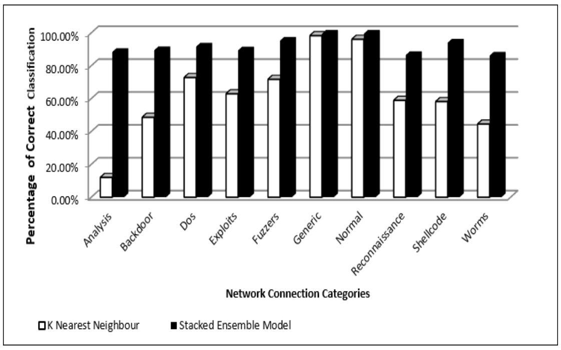


Figure 5. Classification Performance of K-Nearest Neighbor and Stacked Ensemble Model

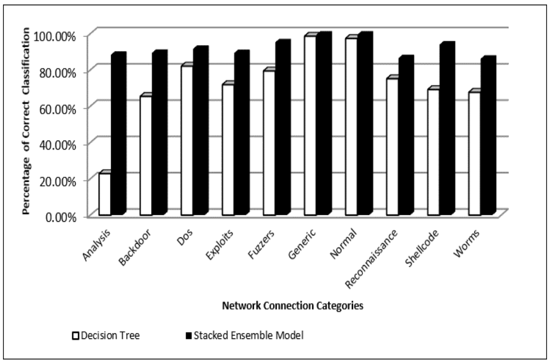


Figure 6. Classification Performance of Decision Tree and Stacked Ensemble Model

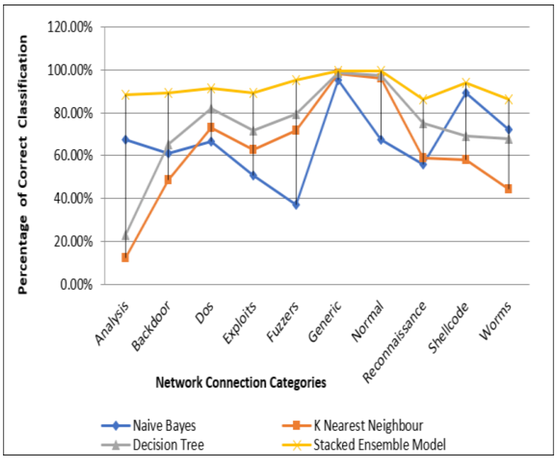


Figure 7. Line Graph of Classification Performance of All the Models

Table 3. Binary (Attacks and Normal Label) Evaluation of the Base Models and MMT stacked Ensemble Model

	NB	KNN	DT	MMT
Classification Accuracy (%)	70.20	82.05	86.77	**97.93**
False Alarm Rate (%)	4.37	2.46	1.62	**0.22**

6. Conclusion

This work investigates the possibility of using stacked ensemble to improve classification performance of a network intrusion detection, we implement three base models, their predictions are used to implement a stacked ensemble intrusion detection system; the Stacked Ensemble system records improved multi-class and binary classification performance over the three base models. Performance comparison among the base models shows Decision Tree model has the highest classification accuracy in seven out of the ten network connection categories, followed by Naive Bayes model with the remaining three out of the ten network connection categories. Despite the fact that KNN model has the least classification accuracy among the three base models, it records a higher classification accuracy than Naive Bayes in six out of the ten network connection categories.

The results in Table 2 and Table 3 confirm the ability of stacked ensemble model to record improve performance over base models. The MMT stacked Ensemble model records higher classification accuracy and the lowest false alarm rate more than the highest accuracy and less than lowest false alarm rate recorded by any of the three base models.

7. References

[1] D. Landes, E., Schumann, S. Schlottke"Identifying Suspicious Activities in company Networks through Data Mining and Visualization. In: P.Rausch, A.F. Sheta and A. Ayesh (eds.) Business intelligence and Performance Management, Springer 2013, pp. 75-90.

[2] S. Pontarelli, G., Bianchi, S., Teofili, "Traffic-Aware Design of a High-Speed FPGA NetworkIntrusion Detection System. IEEE Transactions on Computers" 2013, 62(11), 2322-2334

[3] R. E. Schapire, "The strength of weak learnability. Mach Learn" 1990, 5(2):197–227.

[4] M. Aydin, M. A. Ali, H. Zaim, and G. Ceylan. "An hybrid Intrusion Detection System Designforcomputer network security", Computers and Electrical Engineering, 2009 p 517-526

[5] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection". J. Mach. Learn. Res. 3, 2003, pp1157–1182

[7] M. Gudadhe, P.Prasad and K. Wankhade "A New Data Mining Based Network Intrusion Detection model". 2010, pp 731-735. 10.1109/ICCCT.2010.5640375.

[8] M. Panda and M. R. Patra "Belief Network with Genetic Local Search for Detecting Network Intrusions", international journal of secure digital information age; 2009, 1(1):34-44

[9] P. Garcia-Teodoro, J. Diaz-Verdeio, G. Macia-Fernandez, E. Vazquz, Anomaly-based Network Intrusion Detection: Techniques, systems and challenges Journal of Computers and Security, Volume 28 Issue 1-2, February, 2009 Pages 18-28

[10] B. K. Alese, "Alert Correlation in Intrusion Detection System, International Journal of Physical Sciences" 2006, 1:(1):59-62.

[11] Li Yang. A Lightweight Intrusion Detection Model Based on Feature Selection and Maximum Entropy Model. International Conference on Communication Technology (ICCT '06), 2006,1-4

[12] N. and J.Slay.: The Evaluation of Network Anomaly Detection Systems: Statistical Analysis of the UNSW-NB15 Data set and the Comparison with the KDD99 Data set", in Information Security Journal: A Global Perspective, 2006, 25 :18–31.

[13] M. K. Lahre, M. T. Dhar, D. Suresh, K. Kashyap and P. Agrawal "Analyze different Approaches for IDS Using KDD99 Dataset," International Journal on Recent and Innovation Trends in Computing and Communication, 2013, 1(8): 645–651

[14] A. O. Adetunmbi, S. Zhiwei, S. Zhongzhi and O. S. Adewale"Network Anomalous Intrusion Detection using Fuzzy-Bayes, in International Federation for Information Processing (IFIP)," Boston: Springer 2006, 228: 525–530.

[15] G. Wang, J. Hu, Q. Zhang, L. Xianquan and J. Zhou "Granular Computing Based Data Mining in the Views of Rough Set and Fuzzy Set. Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation".2009, pp 67. 10.1109/GRC.2008. 4664791..

[16] H. Wolpert David "Stacked Generalization, journal of Neural Networks", 1992, 5: 241-259

[17] L. Todorovski and S. Dzeroski "Combining Multiple Models with Meta Decision Trees. Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery, Springer, Berlin, Germany, 1997 pp 54-64

[18] J. R. Quinlan, Book Review: C4.5: Programs for Machine Learning. San Francisco: Morgan Kaufmann. Publishers, Inc..2003

[19] Choudhury S. & Bhowal A., (2015) Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection, Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference,

[20] Tama B. and Rhee K. H., A combination of PSO-based feature selection and tree-based classifiers ensemble for intrusion detection systems, in Advances in Computer Science and Ubiquitous Computing, (2015) 489–495.

[21] I. Syarif, E. Zaluska, A. Prugel-Bennett,G. Wills Bagging, Boosting and Stacking Application to Intrusion Detection. In: Machine learning and data mining in pattern recognition. Springer;.2012, p. 593–60